Chapter 9
CONCLUSIONS AND FUTURE WORK

This thesis begins with a survey of the foundations of quantitative genetics and evolutionary theory in the twentieth century. Ideally, the results of my work presented in this document would directly extend or otherwise complement these classic ideas. They do not. At best, they present some cases that I feel the classic theories do not currently handle well. Their importance to the field will be in the eye of the beholder.

Specifically, my thesis work consists of various ways of looking at the same problem: how the connectivity of genetic networks affects phenotypes in both the short and the long term, and whether or not we can make generalizations from these data. I use certain non-substitutional classes of mutations as examples that lead to interesting evolutionary dynamics, specifically when considering certain familiar models of genetic interactions (e.g. metabolic coupling). I offer some conclusions below regarding the significance of each line of inquiry.

9.1 STRs, Waddington, and the shifting-balance theory

In the first part of this thesis, I focus on a series of anecdotes regarding phenotypic effects of variation in specific STRs. In Chapter 4 I make a case why I expect other STRs to show qualitatively similar behavior The reasoning behind this case stands to be re-emphasized: we should expect evolution to work differently when there is large among-locus variation in mutation rates. In Chapter 4, I argue that, all else being equal, we expect to observe that STRs and functional elements with high mutation rates have above-average epistatic connectivity relative to other loci. This is for evolutionary rather than mechanistic reasons. Specifically, if it is possible to cross a fitness valley

due to a relatively fast STR mutation, then we expect that the fast mutation will be the path taken (Figure 9.1A). I have not treated this intuition with the attention I believe it deserves, but I hope that these STR anecdotes provide a justification for why the problem is interesting. In the summer of 2015, I shared some of these ideas with Christoph Adami, who expressed interest in applying multiple mutation rates in his evolutionary algorithm research program. Hopefully this work will lead to new insights.

One question which I have not discussed is why STRs are located in functional regions in the first place; theory predicts, after all, that selection should minimize the mutation rate at non-neutral loci [216]. This is not entirely inconsistent with what we know; for instance, the *ELF3* STR is not even conserved among the Brassicas, for all its phenotypic relevance in *A. thaliana*, and Figure 9.1A shows a lower equilibrium fitness with a fast mutation rate. Also, in simple models of sign epistasis with two loci, such as are generally used for studying compensatory evolution [347], the convergence to peak fitness is actually slowed down if the two loci have different mutation rates (while maintaining the same overall average rate across loci, Figure 9.1B). Thus, at equilibrium, we would expect generally low mutation rates (e.g. removing STRs from functional regions), and convergence of mutation rates across loci, to maximize fitness and adaptability. If on the other hand equilibrium situations are rare, then perhaps diverse mutation rates would be maintained. Some recent work has suggested that equilibria may be the exception rather than the rule at the level of populations [22]. How to reconcile these dynamics, the dominant theory, and the observed prevalence of STRs remains a question open for speculation.

Chapters 2, 4, 5, and 6 lay out a specific example of the genetics of STRs, the polyQ-encoding *ELF3* STR, as a case with interesting behavior. This STR seems to follow no particular expectation: its genotype-phenotype function is nonlinear and does not generalize across backgrounds, and its genotype is highly variable. Local adaptation appears to act strongly upon this locus, potentially due to its gene product's promiscuous interactions in genetic networks. In chapter 5, I go so far as to suggest that *ELF3* is an

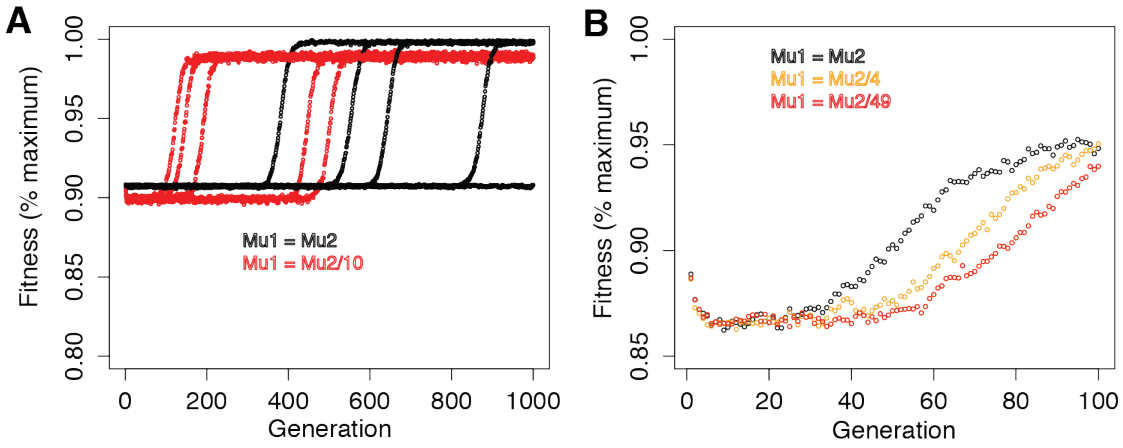

Figure 9.1: Effects of mutation rate variation on simple evolutionary dynamics. (A): Simulations illustrating the scheme in Figure 4.2. Five representative simulations each for two schemes: one with two slow-mutating loci and one with one slow and one fast-mutating locus. In all cases Mu1 = 0.001 (slow). (B): A similar scheme, holding overall average mutation rate constant. A single representative simulation of evolution in a population with two loci with distinct mutation rates, Mu1 and Mu2, which together sum to 0.05 in all cases. All simulations converge to the same fitness value in (B). (A and B): All individuals start with genotype (0,0). Relative to this starting genotype's fitness of 1.0, genotypes (0,1) and (1,0) have fitness 0.5, and genotype (1,1) has fitness 1.1, representing sign epistasis. Population size = 5000 for all simulations.

epistatic hub, similar in its pleiotropic effects to the well-known molecular chaperone Hsp90 [310, 312]. Indeed, in as-yet-unpublished experiments in collaboration with the Davis lab, I was able to show that ELF3 is epistatic to Hsp90 in the expressivity of a particular phenotype (Figure 9.2). This observation suggests that ELF3's pleiotropy could in some cases actually be acting through Hsp90-dependent pathways, or vice versa. In the verbiage of the Hsp90 field, this would place ELF3 in an ideal position to cause stochastic decanalization of phenotypes through mutation [294, 311], allowing *A. thaliana* to explore greater regions of phenotypic space in a heritable fashion. In Chapter 6, we explore a further similarity noted by others [29, 235, 298], under which the activity of ELF3 is (like Hsp90) condition-dependent, being repressed at elevated temperatures. Judging by the remarkable diversity of *ELF3* STR alleles relative to expectations (Chapters 2 and 4), this style of adaptation may be relatively common.

These ruminations do connect back to Wright, albeit indirectly. Wright believed that a stochastic exploration of gene combinations by subdivided populations would be the most efficient way to explore a fitness surface [392, 393]. C.H. Waddington [376] took this idea further, even transgressing against the dogma of Mendelism, and was curious whether phenotypes could explore even greater spaces through Lamarck-like patterns outside of traditional heritability. Very briefly, he reasoned that phenotypic variability or adaptability may itself be a heritable trait. He was able to show that, in cases where variable phenotypes were subject to selection, a specific phenotypic range could be heritably determined ('canalized') when favorable [377]. It would indeed be intriguing if STR variation could be fit into a similar paradigm; it is notable that favorable STR alleles may be stabilized by the acquisition or accumulation of impurity-inducing mutations (Chapter 3, [6]). This series of mutational events may function as one mechanism for Waddington's observations.

The principal question, following this extensive analysis of the *ELF3* STR, is whether or not other STRs will show similar behavior. The answer is not clear. What we can say, from Chapter 4, is that other STRs appear to show a similar excess of observed

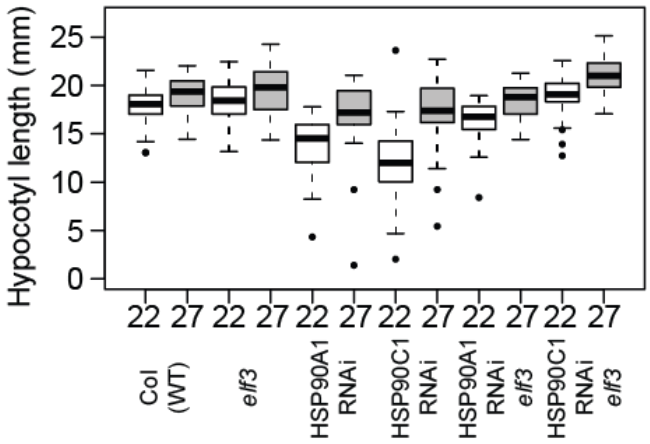

Figure 9.2: *elf3* null mutation is epistatic to Hsp90 inhibition phenotypes. Seedlings were grown for 7d in the dark under various temperatures. 22/27: 22°/27° growth conditions; as shown in Chapter 6, 27° growth conditions represses ELF3 activity and is thus analogous to a *ELF3* knockdown. Hsp90A1 and Hsp90C1 are RNAi lines partially reducing Hsp90 levels [310]. N > 40 for each genotype in each condition. This experiment was repeated with similar results. To be included in a forthcoming publication in collaboration with Ma Zisong and Seth J. Davis.

variation relative to sequence-based expectations. This would lead us to believe that they, like ELF3, show signatures of positive selection. Our model to date for this positive selection is the compensation for other alleles in the background, similar to ELF3, but only experimental tests will show this convincingly.

Next steps

One obvious deficiency of the work presented here is that high-quality STR genotype data are scarce. In Chapter 4, I analyzed a pilot experiment representing some of the first trustworthy high-throughput data generated on STRs. Clearly, if we had data on a more exhaustive sample of STRs, we could come to a better understanding of both the neutral expectation and the effect of selection on STRs. In principle, we could repeat both the association analysis and the selection analysis presented in Chapter 4 with this data, and thereby come to a much better understanding of STR natural variation. We are actively engaged in generating such a dataset.

A further feature of such data would be to generate hypotheses regarding which STRs show potentially interesting evolutionary dynamics and phenotypic associations. Such cases are the best ones for evaluating whether we can generalize our *ELF3* STR results, using genetic and transgenic analysis such as I present in Chapters 2, 3, 5, and 6.

9.2 Gene gain, gene loss, and whole-genome evolutionary trajectories

In the second part of this thesis, I consider genomic evolution at a very high level. Specifically, I mean that I consider only presence and absence of purportedly orthologous sequences within genomes, in the hope that this level of abstraction makes it simple to define acquisition and loss of identical-by-descent sequences across large evolutionary timescales. If we can reliably cluster sequences thus, and the clusters are functionally meaningful, then we can speak of the loss or acquisition of such genes as occurring 'convergently' or 'in parallel' in multiple sequences. For instance, the inde-

pendent acquisition of the same plasmid by two different bacterial cells can be said to be a such an adaptation. We may approach this problem with the hypothesis that some genetic or environmental precondition underlies the adaptation; I shall slightly simplify the terminology by saying that parallel evolution describes the case where there is a genetic precondition, and convergent evolution describes the case where there is an environmental precondition to the adaptation[1] [261].

When parallel evolution is in effect, then the effects of historical contingency are dominant in the form of genetic constraints. This is obviously an instance of epistasis. On the other hand, convergence would seem to indicate the dominance of natural selection in adapting diverse organisms to an environment. This distinction has the nice feature of also corresponding to a traditional ideological divide in evolutionary biology over the relative importance of selection vs. constraint (Chapters 5 and 10 in [117]).

Considering again the gain and loss of orthologous sequences, we may then ask whether there are apparent determinants of such events. If parallel evolution is a meaningful force, then we would expect that we may detect it by examining whether there are dependencies between the gain and loss of a certain sequence and the presence or absence of some other sequence. This we do, using prokaryotic evolution as a case study (where the horizontal acquisition of genetic material is both massive and routine [178], unlike most eukaryotes).

In Chapter 7, I implicitly use the presumption of parallel evolution to detect non-independent evolution between the bacterial *hsp90* gene and other genes with which it has functional links. In this manner, I was able to show that non-independent evolution gives rise to interpretable patterns of gene gain and loss, and collaborators were

[1] This is a non-canonical definition, but it seems more sensible than others I have read, or than lumping the two [348]. These are of course not mutually exclusive, and all instances are necessarily some mixture of the two. Interestingly, in molecular evolution we may exactly discriminate the two, as convergent evolution denotes changes of the same position to the same residue from different residues, whereas parallel evolution means that the same amino acid transition occurs twice independently from the same starting point [411]. Unfortunately, it is difficult to make the same well-defined determination with analog data. In all cases, we must trust statistical inference to differentiate any type of causality from randomness.

able to validate some instances of proteins predicted to be substrates of Hsp90 in *E. coli*. This analysis was far from exhaustive, considering only the coevolution of *hsp90* with all genes, and not the coevolution of all genes with all genes. In this sense, it was limited; I made predictions based on observed evolutionary patterns without having any idea of the prior distribution of such patterns. Nonetheless, predictions were at least somewhat successful, indicating that this sort of ad hoc analysis may be useful for detecting functional correlations.

The problem of generality remained. It was possible that the true *hsp90* associations were the outcome of chance, and that any particular gene would have shown some similar distribution of meaningless coevolutionary patterns. The technique used for the coevolution analysis in Chapter 7 is an explicit test of non-independent evolution. Unfortunately, evolution between genes can be non-independent for many reasons; for instance, it is well-known at this point that larger genomes are more likely to acquire and retain genes by HGT ([60], Figure 9.3). In this case, genes may be correlated in their acquisition or loss simply as an artifact of genome size, rather than because of epistasis per se.[2]

Having shown some success in the small case of *hsp90* interactions with other genes, I considered that we might be able to test this more general hypothesis of genetic historical contingency constraining evolution, while accounting for various confounders. This is Chapter 8. I engineered a relatively crude heuristic for detecting dependencies within genes; this heuristic, however, had the advantage of being relatively conservative, and being compared directly to a null distribution of genes evolving independently. Notably, I did not directly correct for the genome size bias. Even in this relatively conservative case, I found evidence for widespread influence of genome content on gene acquisition by HGT. Furthermore, these individual dependencies led to high overall

[2]Note that this still could be viewed as parallel evolution of the acquisition of genes, in that larger genomes tend to be the ones that gain genes. However, this does not necessarily relate to specific gene-gene interactions such as we describe for *hsp90*.

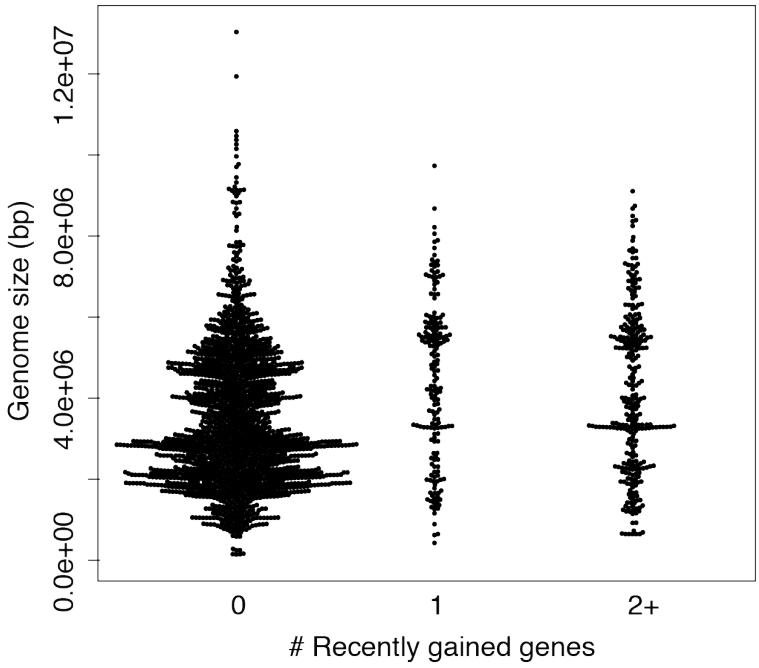

Figure 9.3: Prevalence of recently horizontally acquired genes is weakly correlated with genome size in bacteria. HGT data from [335], sequences were filtered to contain only those genes likely to be inherited horizontally ($\geq$99% identity between sequences from genomes with <97% 16S rRNA identity), and then counted within genomes. Spearman's $\rho = 0.17$ ($p < 10^{-15}$); when correcting for higher gene content of large genomes (HGT/bp), Spearman's $\rho = 0.11$ ($p < 10^{-10}$).

predictability of gene acquisition, highlighting the role of historical contingency in subsequent evolution. Moreover, from my direct ascertainment of genetic dependencies, I was able to show that acquisitions are associated with those specific dependencies, rather than a monolithic effect of taxonomy. This is notable because previous studies have emphasized the tendency of HGT to occur between close relatives [7, 340, 178]. These previous studies suggested that there are higher-order taxonomic determinants constraining gene acquisition, rather than specific genes. My results suggested that at least some of this taxonomic effect can be decomposed and attributed to individual genetic determinants.

Next steps.

I showed in Chapters 7 and 8 that both functional relationships and evolutionary outcomes can be predicted from the coordinated evolution of genes through gains and losses. The prediction methods used in these cases were, however, extremely conservative and rather crude. With more development, and the application of more sophisticated methods from statistical learning, it is likely that the quality of predictions could improve substantially.

However, a larger issue is that in the cases presented so far the outcomes refer largely to gene presence/absence evolution, rather than in phenotypic characters. However, it is the phenotypic characters that are most interesting; thus, it would be very interesting to determine whether phenotypic outcomes (particularly those of medical or technological relevance) can be meaningfully predicted using similar methods. Such predictions could be very valuable in application, beyond simply explicating the genetic basis of phenotypic parallel evolution.

9.3 Some final observations

I once read that mathematics, at its best, is simply 'thinking about things the right way'. I dislike the prescriptive implication of this statement, but I think that it contains a

kernel of truth. Science is a dialectical process by which ideas are found wanting and reformulated. It is through this process that we have decided that we can best know the world. Accordingly, we can know many things about genetics without lifting a finger, provided we are willing to spend some time in rigorous thought considering the data we already have. It is in this light that I have included two quotes at the beginning of this document. The first of these is one of my least favorite witticisms in biology (in a tight race with 'what is true for *E. coli* is true for the elephant'), and the second is a commentary upon the first.

There is an understandable tendency among people (scientists and otherwise) to seize upon simplicity, because we need shorthands to understand the world around us, no matter how poorly they reflect reality. One such shorthand is the idea of additive genetic variance, as discussed in the introduction. Another is the idea of a phylogeny as a bifurcating graph describing the history of life in terms of mitoses. Another is the statement: 'there are two kinds of people in the world: those who divide people into two categories, and those who do not'. Each of these has something to tell us about the world, through its perspective on observed phenomena. Nonetheless, when it is clear from available data that the model is thinking about things the *wrong* way, we might start looking for another model for understanding the world.

Let us consider the case of the heritability of human traits. Much ink has been devoted to discussing the observation that heritability estimates from locus-free estimates such as twin studies consistently estimate that the heritability of traits is much larger than that found by summing the effects of loci discovered by genome-wide association (GWA) analysis (see [223, 80] for examples). There are many competing hypotheses for explaining this observation. Among other explanations these include genetic interactions [418, 249], failure to ascertain and/or tag causal variants [80], rarity of variants [50, 80], and environmental contributors [80]. One additional explanation is simply that that we have not sampled enough human genetic variation to detect all the causal variants at the relevant statistical thresholds, and that an expansion of the existing re-

search program will be sufficient to predict relevant phenotypic variation [301, 375]. As yet, it is premature to judge the relative importance of these hypotheses.

There is some evidence that, by switching to overall summary statistics rather than attributing specific effect sizes to specific loci, the 'missing' heritability is suddenly accounted for by common polymorphisms [398, 397]. These findings have in turn been challenged on various methodological points [182, 184], and debate is ongoing [399]. In essence, these models (using GCTA [398] as an example) assume that contributing loci are additive and population structure is negligible, and evince bias under other conditions. A strong belief in these assumptions will necessarily run into issues discussed in the Introduction. From this difficulty, I would submit that the question is then not whether causal loci are additive and individuals are unrelated, but rather whether their deviance from additivity and unrelatedness is small enough to be acceptable to GCTA's assumptions. It seems there are often reasons to be skeptical that this is so.

Firstly, we know that the replicability of GWA studies is relatively poor, specifically between ethnicities [36]. There are many explanations of this phenomenon, from different allele frequencies to epistasis to study design. However, the balance of evidence indicates that a GWA study in a given population lacks generality [36]. Second, we know that many effects may be contingent not only on genetic background but also environment; specifically, one recent study indicated that the trait of fertility (for which $h^2 \approx 0.5$) has very little consistent heritability explained by common genetic variation across global populations, but that by stratifying by time and geography, different strong genetic effects became detectable in each population by GCTA [367] (fertility is of course a non-canonical trait in current human quantitative genetics). Third, several very recent high-profile studies reported that, among large cohorts of healthy individuals, dozens were found who had homozygous recessive or dominant alleles judged to cause severe or lethal Mendelian disease [85, 44, 248]. In some cases, these mutations were confirmed in healthy individuals in targeted tests (including a homozygous *PRDM9* knockout) [85, 248]. These observations were ascribed to low penetrance of

mutations, for which an obvious hypothesis is epistatic masking, a common observation in model organisms [293]. One prominent geneticist was drawn to comment on one of the (much-hyped) papers on Twitter: "'Superheroes' in human genetics! In flies we call them CPDs, be less self-congratulatory."

If the effects of even such mutations are context-dependent, then it is difficult to support the model of predominant additivity, and moreover it is difficult to trust results based on sophisticated algorithms making such assumptions. It is notable that, in the second case [367], the genotype-by-environment interactions were uncovered using relatively crude ad hoc procedures to separate individuals into 'treatments' before analysis. It is remarkable how the GCTA software [398], a major technical advance in genetics, could in this instance only be applied in this context by an adaptable user who subverts it by creating a new model, using their intuition of one of the oldest ideas in genetics, the reaction norm [391].

My own experience in science has been similar. I found it more viable where possible to develop ad hoc methods which more faithfully represent a given problem, rather than to attempt to fit my problems into the context of a pre-existing method rich in biologically implausible assumptions. Examples of both types of study are readily available in this dissertation. I will let the reader judge the result, and simply submit that Brenner's hierarchy, as quoted, should probably be inverted.